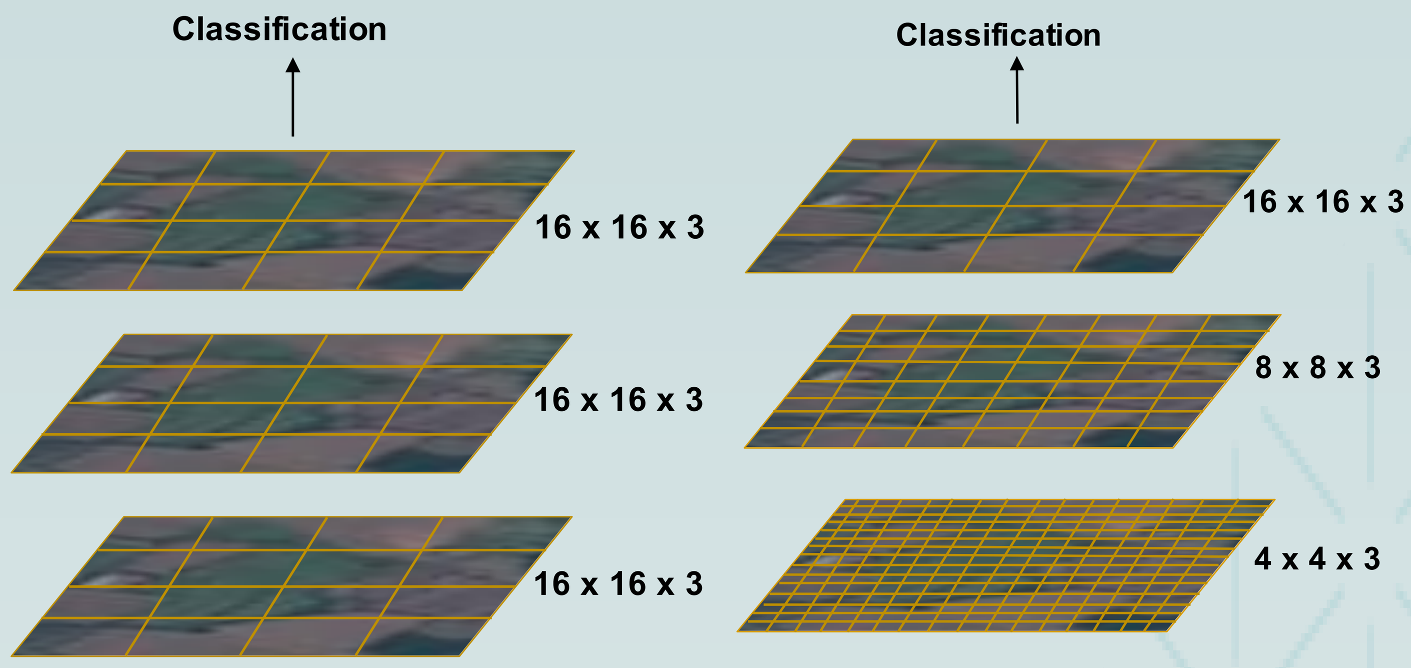
Explainable Land Use and Land Cover Classification Based on Transformers Using Satellite Imagery

Mehak Khan, Abdul Hanan, Meruyert Kenzhebay, Michele Gazzea, Reza Arghandeh
Department of Computer Science, Electrical Engineering and Mathematical Sciences
Western Norway University of Applied Sciences, Bergen, Norway

Background

- Studying Land Use and Land Cover (LULC) is critical to understanding human-environment interactions and helping with decision-making in fields like agriculture, urban planning, and sustainability.
- Transformer-based models enhance LULC analysis with high accuracy but often require substantial computational resources, posing a significant challenge in the field.
- In addition, there is an issue associated with their complex and black-box nature. Therefore, making them transparent is becoming increasingly essential to understand how decisions are made.



(a) Vision Transformer (ViT) (b) Swin Transformer

Fig. 1: Vision Transformer vs Swin Transformer.

Proposed Framework

- This study proposes an innovative framework that uses optical satellite imagery for LULC classification.
- The proposed framework distinguishes itself by exploiting the advanced capabilities of transformer-based models and reduce their computational demands using Transfer Learning and Fine-tuning.
- A notable aspect of this framework is its focus on explainability using Captum, integrated into its algorithmic architecture.
- This integration of explainability ensures that the framework not only performs its task effectively but also provides insights into the decision-making process of the black-box transformer-based model, making it a robust tool in the field of satellite-based LULC analysis.
- The proposed framework is divided into two main blocks, Transformer-based LULC and Explainability, visually depicted in Fig. 2.

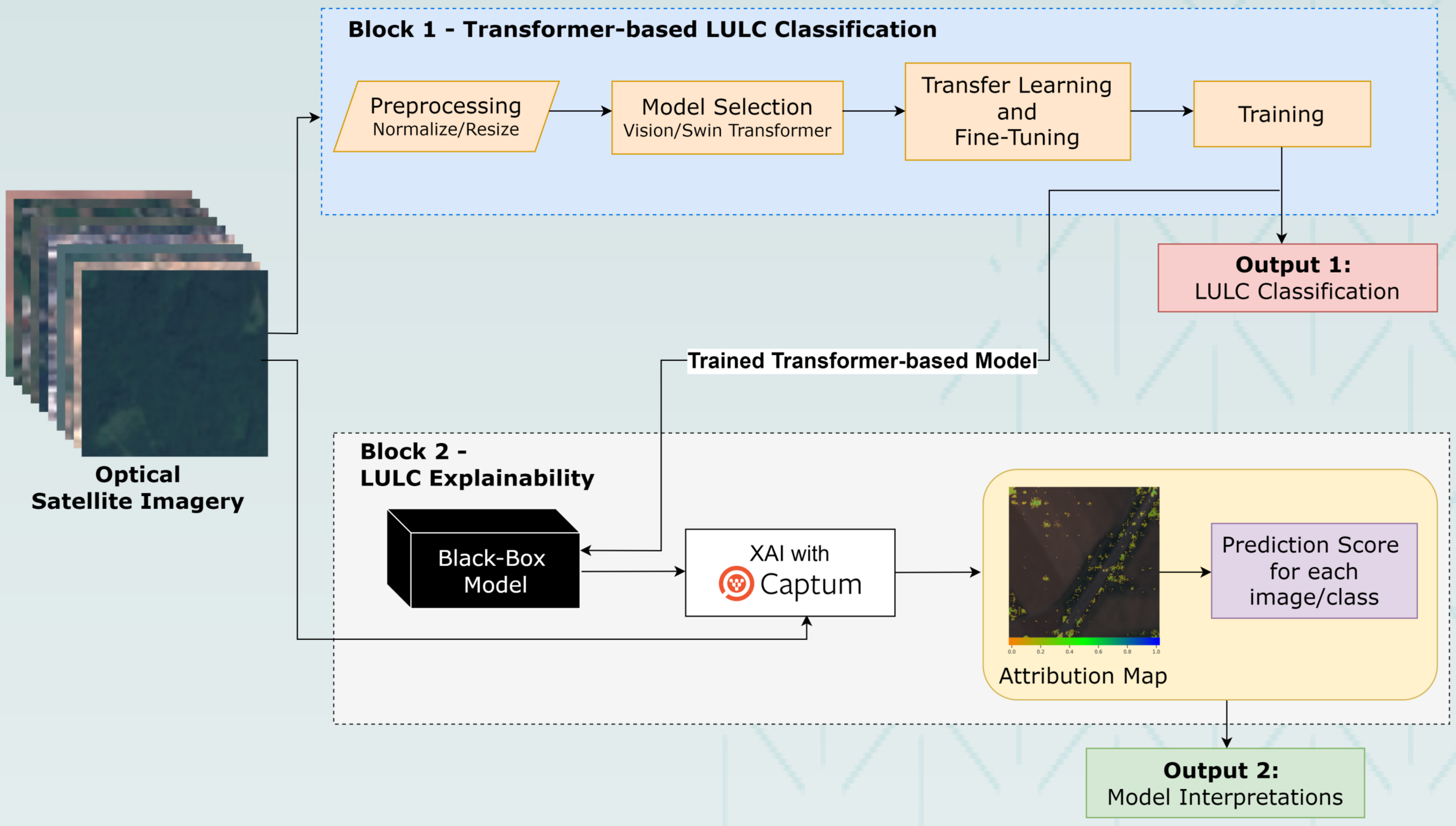


Fig. 2: Overview of the proposed framework.

Experimental Results

- Experimental results demonstrate that transfer learning can significantly improve the performance of transformer-based models for satellite image classification and LULC analysis. Concurrently, strategic fine-tuning can reduce computational costs with only a minimal drop in accuracy, Table 1.
- t-SNE visualizations demonstrated in Fig. 3 shows the effectiveness of our framework. All the models show well-defined clustering of features post-transfer learning, which leads to a clear separation between different LULC classes.
- Fig. 4 highlight the critical role of explainability block in our proposed framework, ensuring fairness, transparency, and trust in AI-based decisions, especially for LULC. Attribution maps support computer vision and satellite image analysis and other earth observation tasks.

Models	Test Loss	Test Accuracy	F1-score	Balanced Accuracy	Computation Time	Parameters
ViT-Base/16 - 12 Blocks						
W/o Transfer Learning	0.5866	79.26	78.95	78.52	32.0 minutes	86,567,656
With Transfer Learning	**0.0417**	**98.67**	**98.67**	**98.60**	31.6 minutes	86,567,656
Unfreeze last 3 Block	0.0625	97.83 (-0.84)	97.84	97.77	21.8 minutes (-9.8)	21,263,616
Unfreeze last 2 Block	0.0693	97.65 (-1.02)	97.64	97.40	**16.1 minutes (-15.5)**	**14,175,744**
ViT-Large/16 - 24 Blocks						
W/o Transfer Learning	0.4305	83.93	83.73	82.98	111.2 minutes	304,326,632
With Transfer Learning	**0.0290**	**99.11**	**99.11**	**99.07**	111.0 minutes	304,326,632
Unfreeze last 3 Block	0.0452	98.50 (-0.61)	98.50	98.39	63.7 minutes (-47.3)	37,788,672
Unfreeze last 2 Block	0.0556	98.20 (-0.91)	98.21	98.06	**56.2 minutes (-54.8)**	**25,192,448**
Swin Transformer-Small/16 - 12 blocks						
W/o Transfer Learning	0.7090	74.91	74.52	73.68	48.9 minutes	49,728,418
With Transfer Learning	**0.0516**	**98.28**	**98.28**	**98.23**	48.7 minutes	49,728,418
Unfreeze last 3 Block	0.0564	97.98 (-0.30)	97.98	97.92	44.2 minutes (-27)	48,723,348
Unfreeze last 2 Block	0.0567	98.17 (-0.11)	98.16	97.90	**41.6 minutes (-37)**	**47,750,664**

Table 1: Performance comparison among Transformer-based models utilizing Transfer Learning and Fine-Tuning.

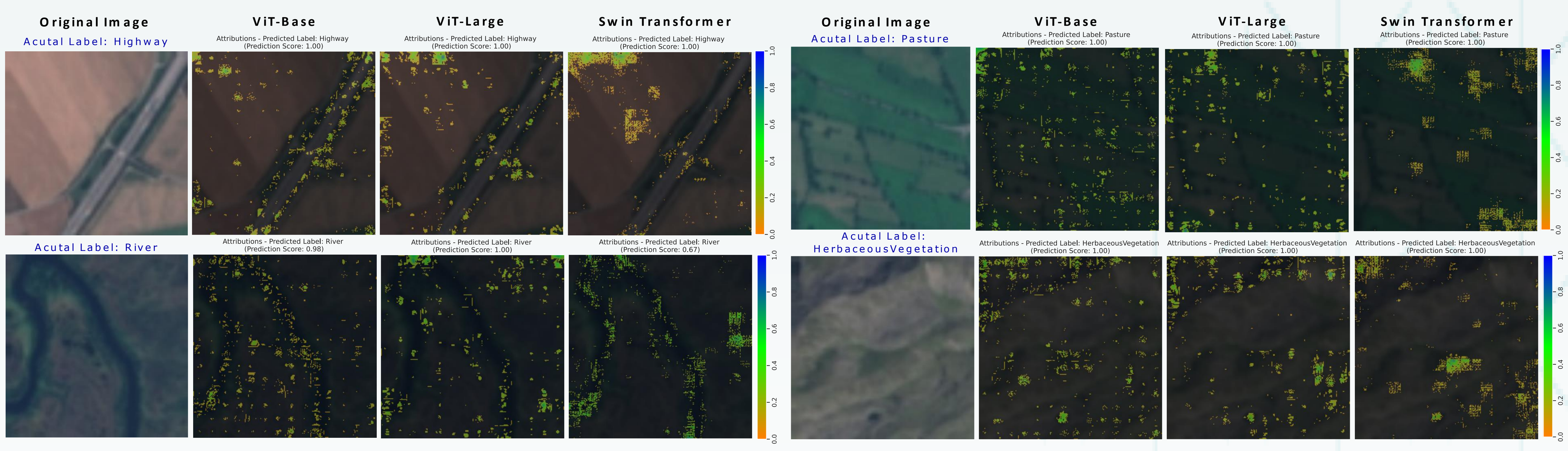
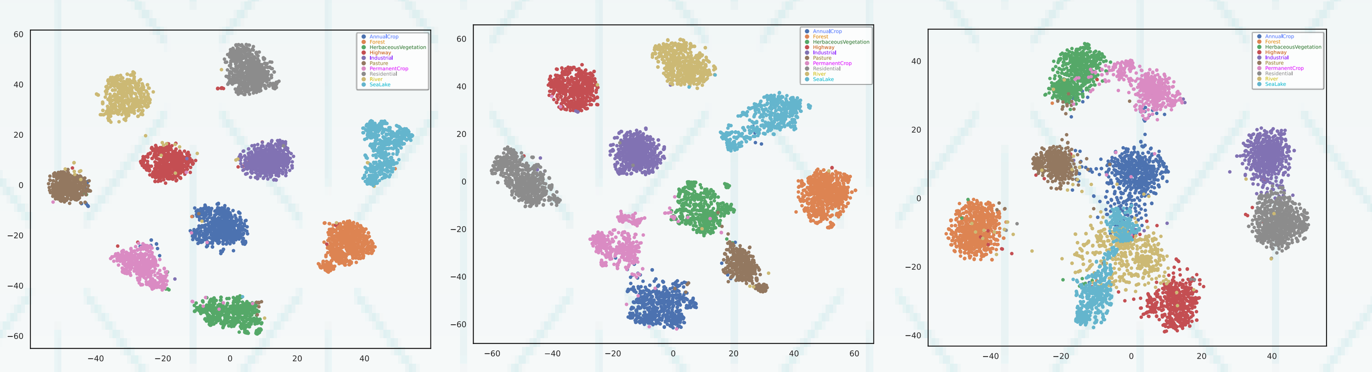


Fig. 4: Attribution Maps representing different LULC Classes across Transformer-Based Models.



(a) ViT-Base (b) ViT-Large (c) Swin Transformer

Fig. 3: t-SNE Visualization: Distinctive patterns in Vision Transformer and Swin Transformer feature spaces.

Western Norway University of Applied Sciences

CI2 CONNECTIVITY, INFORMATION & INTELLIGENCE LAB

Acknowledgement

We thanks our partner company StormGeo and the support from European Space Agency and agency for GridEyeS project.

esa

StormGeo